This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3102965, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

Passive Microwave Brightness Temperature Assimilation to Improve Snow Mass Estimation across Complex Terrain in Pakistan, Afghanistan, and Tajikistan

Jawairia A. Ahmad, Barton A. Forman, Edward H. Bair, and Sujay V. Kumar

***Abstract*—An ensemble Kalman filter is used to assimilate Advanced Microwave Scanning Radiometer-2 (AMSR2) observations of passive microwave (PMW) brightness temperatures (spectral differences, ΔT_b) into land surface model estimates of snow mass over northwestern high mountain Asia (HMA). Trained support vector machines (SVMs) serve as the observation operator and map the geophysical modeled variables into ΔT_b space within the assimilation framework. Evaluation of the assimilation routine is carried out through comparison of assimilated snow mass estimates with an in situ dataset. The assimilation framework helps improve the land surface model estimates through PMW ΔT_b assimilation, particularly in terms of decreasing the domain-wide bias. The assimilation framework proved more effective during the (dry) snow accumulation season and decreased the bias and RMSE in snow mass estimates at 76% and 58% of the comparative pixels, respectively. During the snow ablation season, the PMW brightness temperature signal contained less information related to snow mass due to the presence of other concurrent geophysical features that effectively serve as noise during the snow mass update. The utilization of PMW ΔT_b for accurate snow mass estimation in complex terrain such as HMA is dependent on a multitude of factors for optimal results; however, it does add utility to the land surface model if the relevant pitfalls are taken into consideration prior to the state variable update.**

***Index Terms*—passive microwave, brightness temperature, land surface modeling, hydrology, snow, high mountain Asia, NASA Land Information System**

I. Introduction

SNOW water equivalent (SWE) represents the amount of water obtained if the snowpack was converted into liquid water. SWE is an important variable in the context of local and regional hydrology. The equivalent amount of water contained within the snowpack influences runoff in rivers downstream during the snow ablation period. Rivers that originate in the mountain ranges of high mountain Asia (HMA) depend on the snow and glacier melt, to varying degrees, for their runoff [1], [2]. Consequently, the population residing in those river basins depends significantly on the runoff generated [3], [4], which in turn is dependent on the snow and ice melting patterns upstream. However, there is considerable uncertainty regarding the spatial and temporal variation of snow in HMA. This uncertainty must be minimized in order to better understand the current snow mass (and corresponding runoff) variability and the resulting implications for regional freshwater access.

Various methods, utilizing different ranges of the electromagnetic spectrum, have been used to estimate SWE in the past. SWE estimation algorithms have traditionally used the difference between brightness temperatures observed at lower (e.g., 10.7 or 18.7 GHz) versus higher frequencies (e.g., 36 GHz) to retrieve SWE information [5]. Radiation emitted at a higher frequency from the underlying soil is preferentially scattered by the dry, overlying snowpack as compared to radiation emitted at a lower frequency at the same polarization [6]. Brightness temperature, T_b, is a measure of the radiation emitted by a surface while brightness temperature spectral difference, ΔT_b, is the difference between T_b observed at different frequencies at a given polarization. More specifically, the Raleigh-Jean approximation for microwave radiation defines T_b as the product of emissivity (a dimensionless surface property) and physical temperature of the microwave emitting surface [7]. Satellite-based radiometers observe the brightness temperatures of the underlying surface, which are subsequently used to extract information about relevant geophysical variables such as snow water equivalent [8]. Alternatively, reconstruction methods utilize the snowmelt generated by the snowpack to estimate the total integrated SWE [9]. Molotch and Margulis [10] used a snowmelt simulation model and visible imagery from Landsat Enhanced Thematic Mapper (ETM+), Moderate Resolution Imaging Spectroradiometer (MODIS), and Advanced Very High Resolution Radiometer (AVHRR) to estimate the amount of SWE accumulation. While providing useful information, reconstruction techniques are dependent on the complete melt-out of the snowpack (end of the snow season) before the snow depletion curves can be developed.

Several previous studies have endeavored to estimate SWE in HMA. Kirkham et al. [11] attempted SWE estimation in the Langtang Valley catchment in Nepal by amalgamating numerical modeling techniques with in situ data. Bair et al. [12] attempted SWE prediction in the watersheds of Afghanistan using machine learning algorithms and physiographic predictors. Although encouraging results have been achieved through various schemes at the basin or catchment level [11], [13], there is a need to develop a technique that can be consistently

J. Ahmad and B. Forman are with the Department of Civil and Environmental Engineering, University of Maryland, College Park, MD, USA (jahmad@umd.edu; baforman@umd.edu)

E. Bair is with the Earth Research Institute, University of California, Santa Barbara, CA, USA (nbair@eri.ucsb.edu)

S. Kumar is with the Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA (sujay.v.kumar@nasa.gov)

applied over the whole data-scarce HMA domain, including the Hindu Kush and Himalaya mountain ranges in the west and center as well as the monsoon-dominated regions in the center and east of HMA. In this study, we explore the utilization of data assimilation to formulate a consistent technique for SWE estimation over the whole HMA domain.

Data assimilation (DA) is the integration of observed data into model estimates. Synthetic experiments carried out by Kwon et al. [14] to study the applicability of PMW ΔT_b assimilation to improve SWE estimation across HMA showed that the assimilation framework was effective in improving SWE estimates for SWE depths < 200 mm during dry snowpack conditions. Building upon these synthetic tests, this study utilizes real-world PMW ΔT_b observations from the Advanced Microwave Scanning Radiometer-2 (AMSR2) in an attempt to improve snow mass estimates across HMA.

II. Study Domain

High mountain Asia (HMA) is generally defined as the high elevation region within the Asian continent that spans over eight countries–Tajikistan, Afghanistan, Pakistan, India, China, Nepal, Bhutan, and Bangladesh–and three main mountain ranges–Hindu Kush, Karakorum, and Himalaya (Fig. 1).

According to the Sturm and Holmgren classification, the snow in HMA is primarily composed of prairie and ephemeral snow types along with the presence of alpine and maritime snow near the glacier zones located at the border between Pakistan and China [15]. Hammond et al. [16] observed that low snow zones coincide with areas of low elevation and that snow persistence increases with elevation. Several studies have analyzed the loss of snow cover and glacier melt under evolving climatic scenarios in this region [17]–[19]. Heterogeneous trends in seasonal SWE were reported across HMA based on a PMW-based analysis of change in seasonal SWE [20]. Changes in seasonal snow affect the downstream runoff, especially for the Indus Basin [21]. Cryospheric monitoring of this area is as important as it is difficult due to the harsh climate and inaccessibility of the mountainous regions.

Remote sensing in HMA is a complex process primarily due to the high spatial variability in elevation, relatively coarse resolution of available satellite data, the relatively consistent presence of clouds, and a general lack of ground-based measurements for model validation and evaluation purposes. In such situations, data assimilation (DA) aids in understanding and improving the estimation of the various geophysical variables such as SWE. In this study, the northwestern part of HMA spanning Pakistan, Afghanistan, and Tajikistan is examined. Data assimilation (DA) of PMW ΔT_b is attempted to improve snow mass estimates in this region. Further details regarding the data assimilation framework are provided in Section III, the results of which employ in situ data during evaluation of the DA results.

III. Assimilation framework

A. Land Information System and Noah-MP

The NASA Land Information System (LIS) is a high-performance computing framework used for land surface

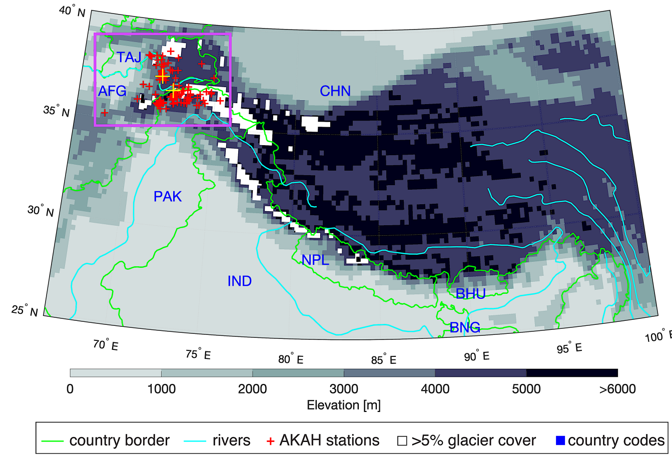

Fig. 1: High mountain Asia comprises multiple countries (country codes in blue text). Glacier cover is computed by up-scaling a binary mask developed using the glacier outlines from the Global Land Ice Measurements from Space (GLIMS) database [22]. The red markers indicate the AKAH stations while the yellow crosses locate the test sites #1 and #2 discussed in Section-V. The purple box demarcates the study domain examined in this study.

modeling and data assimilation [23]. The Noah multi-parameterization (MP) version-3.6 [24]–[26] land surface model is run within the LIS version-7.2 framework to model the land surface conditions over northwestern HMA. The multi-layer snowpack modeling approach implemented in Noah-MP provides a reasonable representation of the interaction between individual snow layers and the geophysical processes inherent in deep snowpacks [25]. Noah-MP simulated geophysical variables were used as input data for support vector machine (SVM) training (Section-III-B) and then as input variables during the prediction of brightness temperature spectral differences using the well-trained SVMs as part of the analysis update aimed at improving the modeled SWE estimates.

Table I highlights the Noah-MP parameter options used in this study. The Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA-2) meteorological forcings were used as boundary conditions to Noah-MP [27]. The Land Data Toolkit [28] was used to preprocess the ancillary data sets on a 0.25° equidistant cylindrical grid. A spin-up period of 4 years starting in May 2012 and ending in October 2016 helped establish suitable initial conditions for the terrestrial hydrologic cycle.

B. Support Vector Machine Regression

Support vector machine (SVM) is a supervised machine learning algorithm developed by Vapnik et al. [39]. In this study, the SVM regression algorithm is used. The SVM framework is composed of two phases: i) training, and ii) prediction. In the *training phase*, known inputs and outputs are used to train support vectors, i.e., define support vectors based on the input data and assign appropriate weights to each support vector [40]. During the *prediction phase*, the trained

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3102965, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

Model Components	**Selected Inputs or Parameterizations**
Elevation, slope, and aspect	SRTM30-v2.0 [29]
Landcover	MODIS (IGBPNCEP) [30]
Maximum albedo	National Centers for Environmental Prediction [31]
Greenness	National Centers for Environmental Prediction [32]
Vegetation	Dynamic vegetation option
Canopy stomatal resistance	Ball-Berry method [33]
Runoff and groundwater	Simple groundwater model, SIMGM [34]
Supercooled liquid water and frozen soil permeability	NY06 [35]
Surface-layer drag coefficient	General Monin-Obukhov similarity theory [36]
Snow surface albedo	Biosphere-Atmosphere Transfer Scheme [37]
Partitioning of rain and snowfall	Jordan91 [38]
Snow and soil temperature	Semi-implicit option
Lower boundary of soil temperature	Noah native option
Meteorological boundary conditions	MERRA-2 [27]

TABLE I: Selection of model components in Noah-MP as implemented within LIS.

SVM input data		
Noah-MP modeled geophysical variable	**Symbol**	**Unit**
Snow water equivalent	SWE	dm
Snow density	ρ_{snow}	dg/m^3
Snow liquid water content	SLWC	mm
Top-layer snow temperature	ST	cK
SVM target data		
Tb spectral difference (frequency and polarization)	**Symbol**	**Unit**
$Tb_{10.7\text{GHz, vert. pol.}}$ - $Tb_{36.5\text{GHz, vert. pol.}}$	ΔT_b $_{10.7\text{V-}36.5\text{V}}$	K
$Tb_{10.7\text{GHz, horz. pol.}}$ - $Tb_{36.5\text{GHz, horz. pol.}}$	ΔT_b $_{10.7\text{H-}36.5\text{H}}$	K
$Tb_{18.7\text{GHz, vert. pol.}}$ - $Tb_{36.5\text{GHz, vert. pol.}}$	ΔT_b $_{18.7\text{V-}36.5\text{V}}$	K
$Tb_{18.7\text{GHz, horz. pol.}}$ - $Tb_{36.5\text{GHz, horz. pol.}}$	ΔT_b $_{18.7\text{H-}36.5\text{H}}$	K

TABLE II: List of SVM input and target training data. Separate SVMs are trained to predict each of the individual ΔT_b combinations.

SVMs are then used to predict ΔT_b using independent input data.

Machine learning techniques such as SVM regression have a unique advantage over regular regression and radiative transfer modeling, i.e., these techniques do not require the explicit representation of each background process via state variables. With respect to radiative transfer models (RTMs) more specifically, RTMs often require inputs that global land surface models cannot provide. That is, global land surface models often lack the fidelity as required by the RTMs, hence, SVM regression provides an alternative approach that enables the use of all of the available information (e.g., soils information, vegetation information) beyond that provided by the snow model and corresponding snow model states. In other words, the input and target data space employed by the SVM regression algorithm does not require the explicit modeling or inclusion of each individual state and parameter [41]. Compared to linear regression techniques, SVM is able to represent the non-linear complexity within the relationship between the input and target data that linear regression techniques cannot. Furthermore, a comparative study of SVM and artificial neural network (ANN) predictions of brightness temperature showed that the SVM outperformed the ANN by capturing more of the high-frequency variability, and in turn, exhibiting a lower RMSE and a higher anomaly correlation coefficient. For places such as HMA, where previous (accurate) knowledge of snow mass is limited, SVM regression serves as a reasonable modeling technique that maps the geophysical states into the ΔT_b space. However, some, but not all, of the relevant processes (and state variables) are implicitly represented through each trained SVM based on the relationship between the input and target data. The fidelity provided by machine learning algorithms make them suitable for use in areas where information about each state variable is lacking. In addition, the SVMs trained at each pixel implicitly represent the local spatiotemporal features such as topography and regional climatology.

Table II presents the SVM input and target data specifications. Based on their sensitivity to ΔT_b [42], four Noah-MP modeled geophysical variables (SWE, snow liquid water content, snow density, and top-layer snow temperature) were selected and used as input data for SVM training and prediction. All the Noah-MP variables were converted into units with similar orders of magnitude in order to better produce appropriate SV weights (third column in Table II). The rescaling via simple unit conversion helps to maintain better consistency between all of the SVM input signals relevant to PMW remote sensing of snow.

The SVM training targets consisted of four different frequency-polarization combinations of the PMW T_b observed by Advanced Microwave Scanning Radiometer-EOS (AMSR-E). PMW T_b observed by the AMSR-E instrument at three different frequencies (i.e., 10.7 GHz, 18.7 GHz, and 36.5 GHz) and two polarizations (horizontal and vertical) were used, Table II. These ΔT_b combinations have been used in numerous studies for SWE estimation [8] [14] [43]. AMSR-E provides T_b observations from 2002-2011 while the AMSR2 T_b observations are available from 2012-present. SVMs were trained using AMSR-E ΔT_b while the AMSR2ΔT_b were assimilated during the DA run. Transferability from AMSR-E to AMSR2 ΔT_b was analyzed (results not shown here) and it was concluded that the AMSR-E ΔT_b trained SVMs were able to predict the AMSR2 ΔT_b climatology. The use of AMSR-E ΔT_b for training and AMSR2 ΔT_b for prediction also provides the necessary independence between the training and prediction datasets.

Fig. 2(a-f) presents the AMSR2 observed T_bs at relevant frequencies and polarizations for one day during the snow

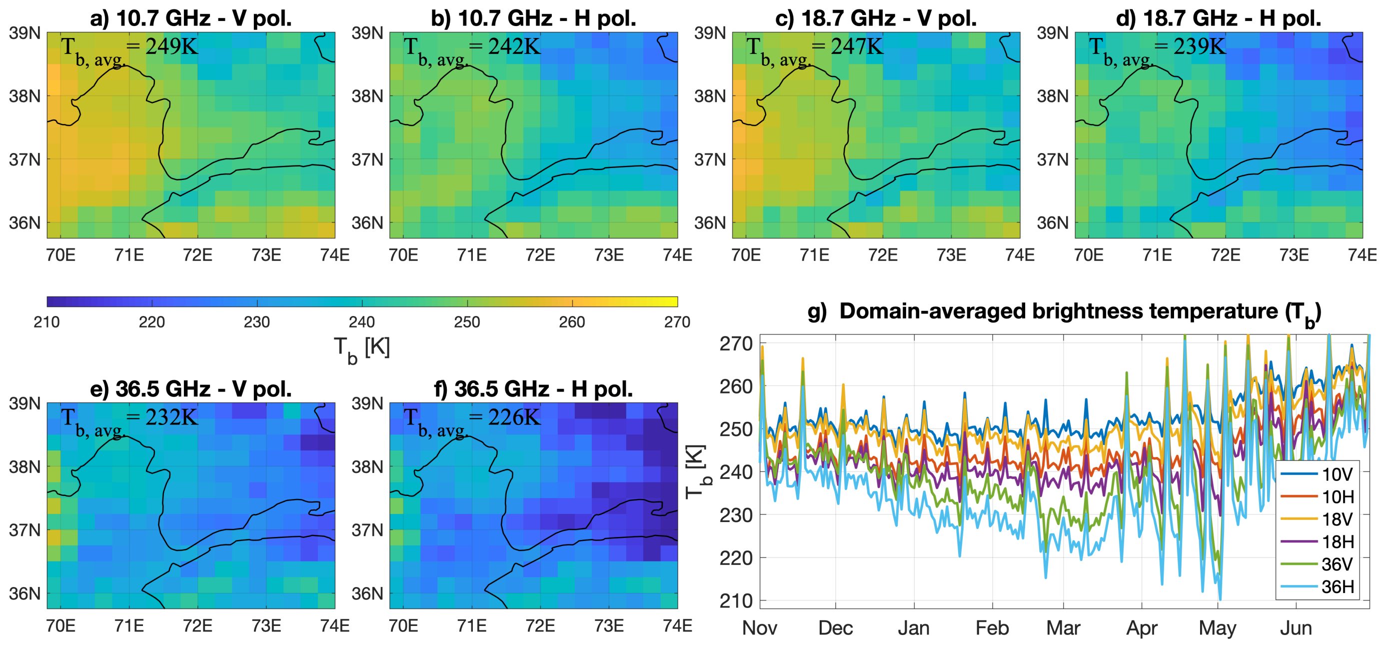

Fig. 2: AMSR2 observed brightness temperatures (T_b) at various frequencies and polarizations across the Pakistan-Afghanistan-Tajikistan sub-domain for February 11, 2017. Subplot (a-f) titles identify the microwave frequency and polarization of the corresponding T_b maps. Subplot g) displays the temporal variation of domain-averaged T_b values for the 2017 snow season. Large differences between the T_bs observed at lower (10.7 and 18.7 GHz) vs. higher (36.5 GHz) frequencies suggest the presence of snow across the area.

season (February 11, 2017). Relatively higher T_b magnitudes were observed across the western part of the sub-domain, which gradually decreased to much lower magnitudes towards the northeast. The differences between the T_bs observed at the lower frequencies (10.7 and 18.7 GHz) versus the higher frequency (36.5 GHz), known as spectral differences, indicate the presence of snow across the region. Fig. 2h displays the seasonal variation in the observed T_bs and highlights the increase in the spectral differences during the snow accumulation months and the subsequent decrease in these differences during the ablation months. Interesting to note is the corresponding increase in the T_b temporal variations (for all frequencies) during the ablation months (April onwards). Only PMW T_bs observed during the nighttime overpass were assimilated to minimize the influence of wet snow conditions. Wet snow not only attenuates the PMW emitted from the Earth's surface but also emits significant PMW radiation itself, as compared to dry snow, thus leading to an increased T_b magnitude observed by the radiometer and possible erroneous estimation of snow [44]. AMSR-E and AMSR2 T_b observations were downloaded from the Japan Aerospace Exploration Agency GCOM website (URL:$https://suzaku.eorc.jaxa.jp/GCOM\_W/data.html$) and mapped to the same 0.25° equidistant cylindrical grid as used by Noah-MP in order to maintain spatial coherence. Further details regarding the SVM theory and training setup are provided in Section-A1.

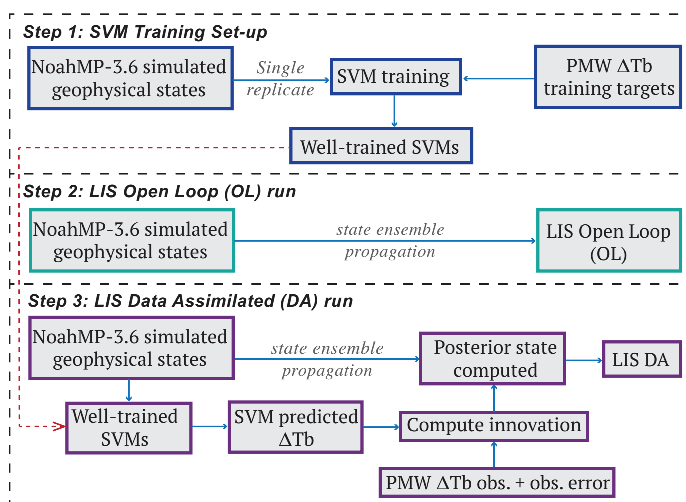

Fig. 3: Three-step approach for the data assimilation experimentation including SVM training, LIS open loop, and LIS data assimilation.

C. Open Loop and Data Assimilation Setup

Fig. 3 illustrates the SVM training, open loop (OL), and data assimilation (DA) experiments. The experiments begin with an open loop (OL) simulation, which is defined as the *model-only* run. Stochastic state and boundary condition (forcing) perturbations are applied to develop an appropriate state ensemble, after which the ensemble replicates are

independently propagated in time using the forward model (i.e., Noah-MP). Forcing perturbation characterizations are the same as those detailed in Table 2 by Kwon et al. [14]. The prognostic state variables (SWE and snow depth) are also perturbed to represent the model structure error via multiplicative perturbations. The ensemble spread implicitly represents the uncertainty in the model predictions due to model structure error and error in the boundary conditions. The OL serves as the benchmark for model-only performance without the benefit of information derived from satellite-observed PMW ΔT_b. The OL and DA runs are propagated through a 20-member ensemble. Since a one-dimensional assimilation framework is used in this study, spatial error correlations were not explicitly considered. However, the general framework outlined here could be expanded to include spatially-correlated errors as part of a follow-on study, but are excluded here in order to maintain a tractable project scope.

The Ensemble Kalman Filter (EnKF) data assimilation algorithm is utilized to assimilate ΔT_b in this study. The DA simulation has two main steps: i) state propagation, and ii) state update. The ensemble development and propagation is similar to the OL, however, the update step requires computation of both the Kalman gain and the innovation by comparing the SVM-based predictions of ΔT_b versus the observed ΔT_b. The ensemble replicate update in the EnKF is computed as:

$$y_{i,t}^{+} = y_{i,t}^{-} + K_t \ Innov_{i,t} \tag{1}$$

$$K_t = C_{y_t z_t}[C_{z_t z_t} + C_{vv}]^{-1} \tag{2}$$

$$Innov_{i,t} = z_t + v_{i,t} - \mathcal{H}(y_{i,t}^{-}) \tag{3}$$

such that i= index of the replicate drawn from a derived probability distribution of size N; $y_{i,t}^{+}$= posterior SWE value of replicate i at time t; $y_{i,t}^{-}$= prior SWE estimate of replicate i at time t; K_t= Kalman gain at time t; z_t = observation at time t; $v_{i,t}$= observation error at time t ($v_{i,t} \sim \mathcal{N}(0, \sigma_{vv}^2)$); $\mathcal{H}(.)$ is the non-linear observation operator that maps the geophysical state space into the corresponding observation (ΔT_b) space; $C_{y_t z_t}$= time-varying cross-covariance matrix between the prior state errors and the predicted observation errors; $C_{z_t z_t}$ = time-varying predicted observation error covariance matrix; C_{vv} = time-invariant observation error covariance matrix; and $Innov_{i,t}$= innovation vector for replicate i at time t. In the EnKF, $C_{y_t z_t}$ and $C_{z_t z_t}$ are approximated from the ensemble sample statistics using the following formulas:

$$C_{y_t z_t} = E[(y_t^{-} - \overline{y_t^{-}})(z_{pred_t} - \overline{z_{pred_t}})^T] \tag{4}$$

$$C_{z_t z_t} = E[(z_{pred_t} - \overline{z_{pred_t}})(z_{pred_t} - \overline{z_{pred_t}})^T] \tag{5}$$

where $E[\cdot]$= expectation operator, y_t^{-}= $\{y_{1,t}^{-}, ..., y_{i,t}^{-}, ..., y_{N,t}^{-}\}$; $\overline{y_t^{-}}$= temporal mean of y_t^{-}; z_{pred_t}= $\{z_{pred_{1,t}}, ..., z_{pred_{i,t}}, ..., z_{pred_{N,t}}\}$; $z_{pred_{i,t}}$= $\mathcal{H}(y_{i,t})$; and $\overline{z_{pred_t}}$= temporal mean of z_{pred_t}.

The EnKF has certain inherent assumptions: i) unbiased, linear forward model, ii) unbiased, linear observation operator, iii) jointly Gaussian and mutually independent observation and model errors, and iv) spatiotemporally uncorrelated errors. However, the EnKF provides certain flexibility such that successful results have been observed for non-linear models and observation operators [14], [43]. Exploiting this flexibility, a non-linear forward model (Noah-MP) and observation operator (SVM) are utilized here.

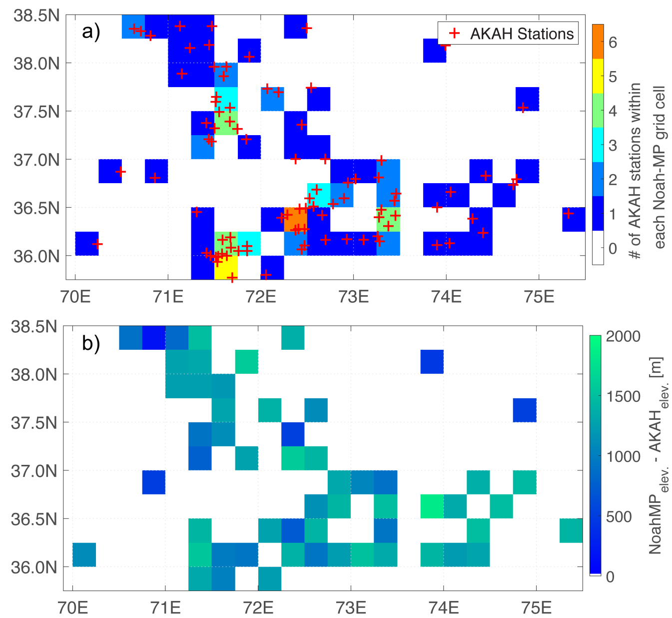

Fig. 4: a) Map of the number of AKAH stations within each Noah-MP grid cell. b) Visual representation of the difference between the Noah-MP grid cell elevation and the average elevation of AKAH stations within the grid cell boundary. AKAH stations within each grid cell are averaged for comparison with the corresponding Noah-MP cell.

IV. Experimental Setup

A. Evaluation Data

Evaluation of the OL and DA results was predicated on the use of in situ snow depth measurements (details in Section-IV-A1). However, due to a lack of temporally-consistent SWE measurements, the in situ snow depth measurements were used as boundary conditions in a separate and independent snow model (SNOWPACK) in order to derive SWE estimates at the station locations (details in Section-IV-A2). SNOWPACK was used to provide temporally-consistent and gap-free timeseries of SWE that could be used as a reasonable proxy for in situ SWE measurements.

1) AKAH Snow Depth: The Aga Khan Agency for Habitat (AKAH) provided in situ snow depth measurements from 93 stations within Tajikistan, Afghanistan, and Pakistan, Fig. 1 and Fig. 4. The in situ measurements were collected at the point-scale whereas the Noah-MP grid size was 0.25° x 0.25° (∼25km x 25km at mid-latitudes).

Fig. 4a presents the location and number of stations per Noah-MP grid cell. Some grid cells contain multiple AKAH stations within the same elevation range. However, several grid cells contain stations which have large elevation differences. For consistency, an average of the station (point-scale) snow depth measurements was computed, wherever applicable,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3102965, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

Type	Update frequency	Assimilation period	Assimilated observations	States used in evaluation
Standard assimilation	$\sim$ daily	Nov 1- Jun 30	AMSR2 ΔT_b	Snow depth, SWE
Data thinning-based assimilation	every 5^{th} day	Nov 1- Jun 30	AMSR2 ΔT_b	Snow depth, SWE
Seasonal assimilation	$\sim$ daily	Nov 1- Mar 22	AMSR2 ΔT_b	Snow depth, SWE

TABLE III: Summary of the LIS-NoahMP assimilation runs described in Section-IV-B.

to compare against the Noah-MP (model-scale) snow depth estimates. For those cells that contained only one station, the station measurements were used directly in comparisons against the model grid cell.

Fig. 4b highlights the difference between the AKAH station elevations and the corresponding Noah-MP grid cell elevations. The AKAH stations have lower average elevation values compared to their corresponding Noah-MP grid cells. The ground stations are generally located at lower elevations (i.e., closer to population) for ease of management and maintenance rather than at the mountain peaks. It is, therefore, acknowledged that the discrepancy in ground station versus grid cell elevation can lead to a potential positive bias in the Noah-MP snow depth estimates.

2) SNOWPACK SWE: SNOWPACK is an open source snow and land surface model focusing on the mass and energy exchange between snow and the atmosphere [45], [46]. Mainly used for avalanche studies, here it is used to model the SWE based on the AKAH snow depth values along with ancillary, downscaled boundary conditions. Due to a dearth of any consistent SWE dataset available in this region, the SNOWPACK model estimates of SWE derived from the AKAH snow depth measurements are assumed to be representative of the ground conditions and are, therefore, utilized in the evaluation of the OL and DA SWE estimates. Previous analysis of SNOWPACK modeled SWE (over Mammoth Mountain, U.S.A.) showed encouraging results relative to in situ SWE measurements when using snow depth (not total precipitation) as the model boundary condition [47]. The quality of the SNOWPACK SWE derived from in situ snow depth measurements used in this study is, however, expected to be less than the SWE modeled using SNOWPACK for Mammoth Mountain, U.S.A by [47]. It is expected that there can be some loss of accuracy as the snow depth measurements are translated into SWE using ancillary data (i.e., downscaled meteorological forcings). In Bair et al. [47], all the energy balance forcings, especially the radiative fluxes and snow albedo, were measured on site. For the AKAH snow depth, on the other hand, downscaled Global Land Data Assimilation System (GLDAS-2, [48]) and Clouds and Earth Radiant Energy System (CERES, [49]) data are used, which can introduce additional error and uncertainty.

The SNOWPACK model forces the modeled snow accumulation to match the in situ snow depth measurements during the accumulation period, while during the ablation period the modeled snow ablation is not updated to match the measured snow depth. Therefore, given reliable daily snow depth measurements, modeled SWE during the snow accumulation season is expected to contain less uncertainty and be closer to the in situ ground measurements, while during the snow ablation season the uncertainty and error likely increases. Comparing the SNOWPACK and AKAH snow depths, the average bias and RMSE were noted to be 4.4 cm and 9.7 cm, respectively. These values are much smaller relative to the average bias and RMSE computed for the Noah-MP OL and DA snow depth estimates (see Table IV). Since daily snow depth data are utilized as boundary conditions to SNOWPACK, AKAH stations that had nearly continuous daily snow depth measurements recorded were selected, i.e., only those stations where greater than 80% of the days in the snow season (Nov-May) had snow depth measurements recorded were used. Therefore, the number of stations included in the SWE evaluation (Section-V-B) is only 52 as compared to the 93 stations used to evaluate snow depth estimates. Further details regarding the SNOWPACK model parameter selection and ancillary forcing inputs is provided in Appendix-A3.

B. *Types of assimilation runs*

Three different types of assimilation runs were conducted: i) standard assimilation, ii) data thinning-based assimilation, and iii) seasonal assimilation (Table III). All of the DA estimates shown here use σ_{vv} = 3 K for the ΔT_b observation error standard deviation. Details of, and motivation for, each assimilation run are provided below.

1) Standard assimilation: The standard assimilation run consists of near-daily (based on AMSR2 overpass frequency) ΔT_b assimilation into the Noah-MP SWE estimates using Equations 1, 2, and 3. This is the standard EnKF implementation run without any modifications.

2) Data thinning-based assimilation: The standard EnKF algorithm does not take rational constraints related to physical processes into account during the update process. For example, it is not likely that a snowpack will melt 100 cm overnight, however, if the ΔT_b observation suggests a 100 cm decrease in snow depth, the computed innovation could result in a sudden and deleterious decrease in the updated snow depth and SWE. Given such a scenario, it is likely that noise or information relevant to other physical features (e.g., soil moisture) in the satellite ΔT_b observations could result in unrealistic snow mass updates. Fig. 5 displays the presence of high frequency noise in the PMW ΔT_b signal, which provides the motivation for conducting data thinning experiments as a means of mitigating some of this noise. The high-frequency variability is categorized as "noise" here in terms of SWE only, i.e., it is related to information in the ΔT_b signal that is *not* relevant to SWE. It does not necessarily represent sensor noise per se. Rather, this is information that masks the snow mass signal that is required for SWE estimation, hence, it acts as noise in terms of SWE estimation specifically. This signal

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3102965, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

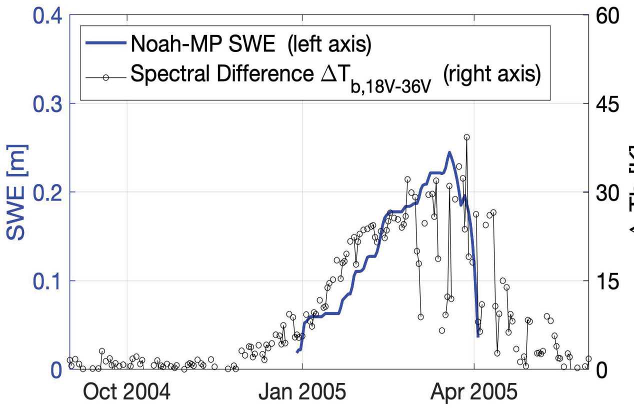

Fig. 5: Modeled SWE estimates via Noah-MP as compared to ΔT_b observations collocated at a point in the HMA study domain. High frequency noise in the PMW ΔT_b observations is translated into noise in the subsequently calculated a-posteriori SWE estimate.

is due to multiple, concurrent processes. The presence of wet snow or even ice layers within the snowpack would affect the electromagnetic response of the snowpack by changing the scattering response of the snow layers [50].

Data thinning is attempted to minimize the occurrence of irrational snow mass updates associated with noise or snow mass-irrelevant information in the ΔT_b observations. Data thinning consists of paring down the available observations such that the assimilation occurs over incremental days. By decreasing the number of assimilation occurrences, it is expected that: i) the possibility of the assimilation of high frequency observation noise into the model estimates will decrease, and ii) the resultant snow depth and SWE timeseries will exhibit a smoother snowpack accumulation and ablation as compared to standard assimilation. An increment of five days was used here such that ΔT_b assimilation occurred on every fifth day. The incremental value was selected based on results yielded by synthetic data thinning experiments carried out by Wang et al. [51] which suggested that an increment of 3-5 days yielded the best results.

3) Seasonal assimilation: Estimation of SWE from PMW radiometry in the presence of wet snow is a complex, ill-posed problem. The influence of wet snow on the PMW radiation emitted from the underlying ground is different as compared to dry snow [52]. The dielectric constant of wet snow is significantly larger than dry snow due to the presence of liquid water within the snowpack. Thus, wet snow not only attenuates the PMW radiation emitted by the underlying ground but also emits its own microwave radiation [44]. The emissivity of dry versus wet snow is significantly different [50]. Therefore, the relationship between SWE and ΔT_b varies for dry versus wet snowpacks [5]. These differences render it difficult to extract only SWE-relevant information from the observed T_b during the (wet) snow ablation period.

The influence of wet snow on the PMW radiation emitted from the underlying ground is different as compared to dry snow. The dielectric constant of wet snow is higher than dry snow due to the presence of liquid water within the snowpack. Thus, wet snow not only attenuates the PMW radiation emitted by the underlying ground but also emits its own microwave radiation [44]. This phenomenon renders it difficult to extract only SWE-relevant information from the observed ΔT_b. Kwon et al. [14] found that SVM-based ΔT_b prediction efficacy was reduced in the presence of snow liquid water content. To overcome this issue, the effect of disabling the assimilation update during significant snow melt (ablation period) was analyzed.

The AKAH stations also provide in situ rainfall data, which was leveraged to better determine a date on which to cease the daily ΔT_b assimilation as part of the seasonal assimilation experiment. The winter rainfall began in March for most of the AKAH stations. The date of occurrence of the first rainfall greater than 5 mm/day for each station was noted. The average (station) median date for snow seasons from 2016 to 2020 was 22 March. Snow cover fraction from the Interactive Multisensor Snow and Ice Mapping System (IMS) snow cover product [53] was also employed as a secondary check. By 22 March, the domain had passed the point of maximum snow-covered area (∼30% of total domain area covered with snow) and had entered the ablation period (∼17% of total domain area still covered with snow). A seasonal run was therefore conducted with standard assimilation until 22 March, after which assimilation was disabled and the model was run in the OL (model-only simulation) configuration. Here we are using a first-order approach to identify the general transition from solid to liquid precipitation (thus indicating a high probability of wet snowpack conditions across the HMA domain).

V. Results

The evaluation results are separated into two parts. In the first part, AKAH snow depth is the geophysical variable used to evaluate the accuracy of the OL versus DA estimates. In the second part, SNOWPACK modeled SWE is used to evaluate the OL and DA SWE estimates. Fig. 4a details the number of AKAH stations within each Noah-MP grid cell used for comparison.

A. Snow depth evaluation

1) Standard Assimilation: The two timeseries in Fig. 6a illustrate the differences between the OL and DA snow depth estimates as compared to the in situ measurements at two test locations (test site locations marked in Fig. 1). For test site #1, the assimilation generally improves the snow depth estimates up until the snow ablation period when the assimilation visibly degrades the predicted estimates, especially during Apr and May, as high frequency noise (in the context of snow mass) present in the AMSR2 ΔT_b signal is assimilated into the modeled snow depth. During the dry snow accumulation months (generally Nov-Feb), the DA estimates are better encapsulated by the standard deviation of the in situ snow depth measurements at stations located within the test grid cell as compared to the OL. Fig. 6a also highlights the presence of a general temporal offset in the snowpack development between the Noah-MP OL simulation and the ground measurements.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3102965, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

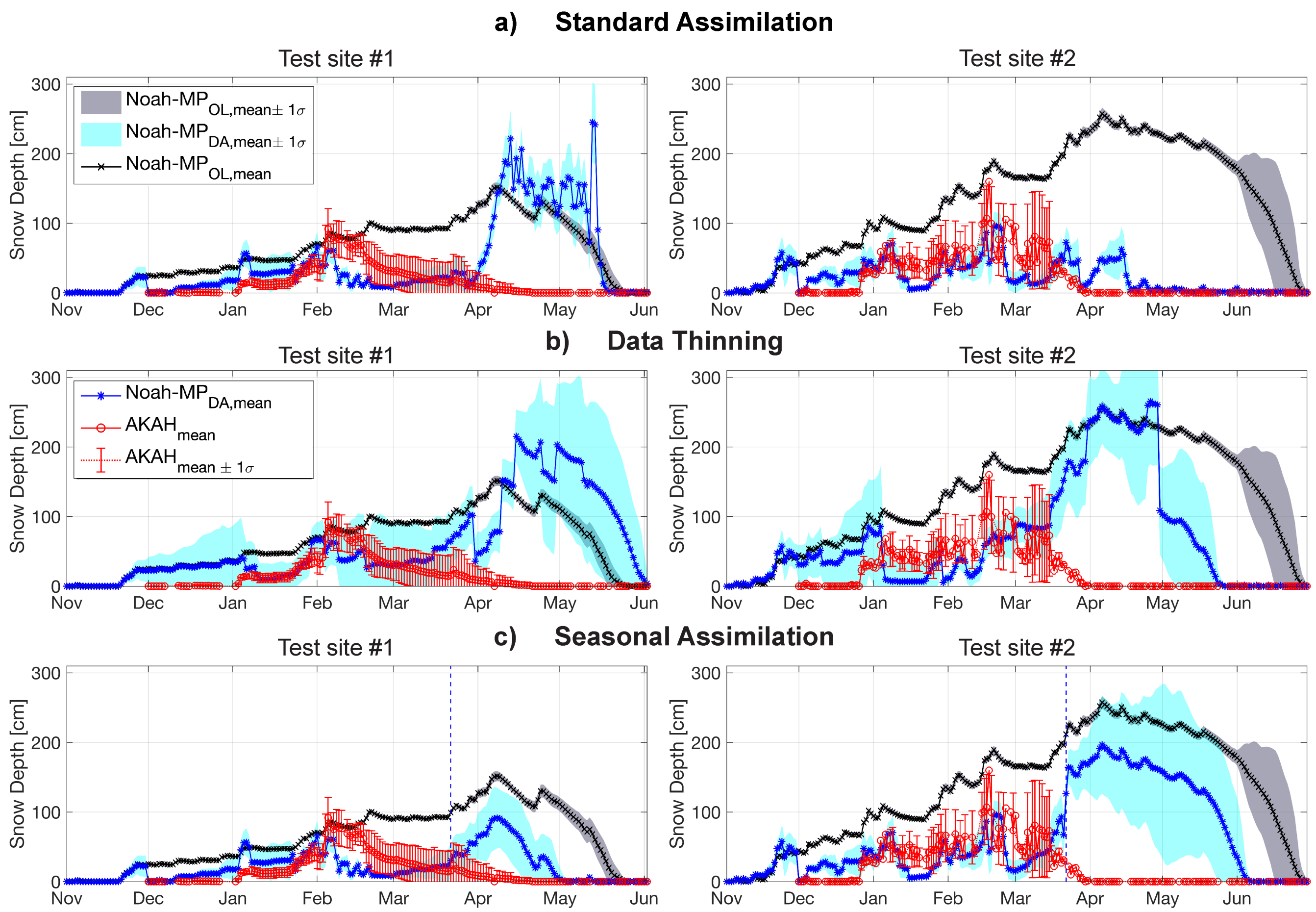

Fig. 6: Comparative timeseries of OL and DA snow depth estimates vs. AKAH measured snow depth for the 2016-2017 snow season. The dashed lines in part c) locate the date (22 March) when assimilation is disabled for seasonal assimilation. The red solid line represents the mean while the red whiskers represent the standard deviation calculated from AKAH station measurements that lie within the Noah-MP grid cell.

The OL model simulation peak occurs in Apr as compared to Feb for the ground stations. For test site #2, ΔT_b assimilation updates and improves the complete melt-out date.

Timeseries for both test sites highlight the increase in uncertainty after assimilation as compared to the OL. The OL ensemble exhibits a restricted dynamic range (ensemble spread) as compared to the DA ensemble estimates. However, the OL ensemble uncertainty increases during the snowpack ablation period. In general, the DA ensemble estimates better encapsulate the AKAH snow depth measurements, especially from Nov-Mar. The increased ensemble uncertainty is apparent in all the DA experiments, i.e., standard assimilation, data thinning, and seasonal assimilation. Therefore, it can be stated that assimilating the PMW ΔT_b observations increases the ensemble uncertainty as the snow mass-related ΔT_b uncertainty is also integrated into the model estimated snow depth uncertainty.

2) Data thinning-based assimilation: In Fig. 6b, the timeseries for test sites #1 and #2 exhibit relatively consistent snow depth changes until the advent of the ablation period. Similar to the standard assimilation run (Fig. 6a test site #1), the snow depth estimates are visibly degraded by assimilation during the snow ablation period. The uncertainty represented by the ensemble spread is much larger for both test sites than for standard assimilation. Observing the two test site timeseries, the data thinning approach yielded smooth sections of the timeseries that underwent sudden, large updates rendering the snowpack development more discontinuous as compared to the standard assimilation results.

3) Seasonal assimilation: The timeseries for test site #1 in Fig. 6a and Fig. 6b showed that the assimilated values and the ground measurements diverge significantly during the snow ablation months. In Fig. 6c, degradation via assimilation is avoided as the assimilation is disabled after 22 March 2017. This also results in a reduced off-set between the ground station measured and the DA predicted snowpack melt-out dates. However, for test site #2, disabling assimilation after 22 March 2017 also results in: i) no opportunity to correct the model estimate through an update (based on ΔT_b assimilation) and ii) an increase in the ensemble uncertainty in a manner similar to the data thinning run. Fig. 6 highlights the location specificity of the test sites' response to the three different assimilation methods. Response of the test grid cells to the three different assimilation approaches varies from location to

location.

Statistic [cm]	Open Loop	Standard Assimilation	Data Thinning	Seasonal Assimilation
Mean Bias	75.4	45.7	63.3	48.6
Median Bias	76.5	32.6	47.5	42.3
Mean RMSE	91.4	76.5	89.4	74.2
Median RMSE	86.2	64.6	74.6	63.4
Mean Unbiased RMSE	49.9	57.6	59.1	52.4
Median Unbiased RMSE	50.7	48.6	57.4	45.5

TABLE IV: Domain-average and domain-median of bias, RMSE, and unbiased RMSE for OL and DA snow depth estimates as compared to the AKAH snow depth measurements for the 2016-17 snow season. All values are in units of centimeter.

4) Statistical results: Statistics in Table IV are computed using the average snow depth of the stations within each grid cell, thus the total number of cells used for calculation is equal to 55. The smallest domain-averaged bias occurs in the standard assimilation snow depth estimates while the smallest domain-averaged RMSE is observed for the standard and seasonal assimilation runs. The domain-median values similarly show better performance for the standard assimilation runs relative to the other simulations. The domain-averaged bias and RMSE for the data thinning run are much larger than either the standard or seasonal assimilation runs, Table IV. However, all three assimilation runs yield better domain-wide statistical results as compared to the OL, Table IV.

The computed statistics were further analyzed with respect to land cover type and elevation. It was found that the average bias in snow depth estimates for pixels categorized as barren land cover was relatively lower for standard and seasonal assimilation runs (36 cm and 41.5 cm respectively) as compared to the OL and data thinning runs (79 cm and 67 cm respectively). Correlations calculated between pixel elevations and statistics (bias and RMSE) were all; i) negative (i.e., as elevation increases the snow depth estimates improve), and ii) $|R|$ <0.5 in magnitude.

Table V highlights the influence of assimilation on pixels within different elevation ranges. Standard assimilation exhibits the lowest mean bias for both elevation ranges and the lowest mean RMSE for elevation >4000 m. Standard assimilation seems relatively more effective for pixels at elevations <4000 m. The mean bias is reduced from 88.8 cm to 52.2 cm (reduction of 36.6 cm) while the mean RMSE is decreased by 18.5 cm through standard assimilation as compared to the OL. For elevations >4000 m, where deep snow is expected, the difference between the various runs is relatively reduced. The maximum reduction in mean bias is equal to 22.4 cm while the maximum reduction in mean RMSE is equal to 11 cm only. According to the results in Table V, ΔT_b assimilation exhibits greater utility at elevations below 4000 m, specifically in reducing the bias in modeled estimates.

Run Type	Snow depth: Elevation of Noah-MP pixels: Mean Bias [cm]		Mean RMSE [cm]	
	≤4000 m	>4000 m	≤4000 m	>4000 m
Open Loop	88.8	61.5	105.3	76.9
Standard Assimilation	52.2	39.1	86.8	65.9
Data Thinning	75.5	50.6	102.3	76.0
Seasonal Assimilation	54.2	42.8	80.9	67.3

TABLE V: Influence of elevation on ΔT_b assimilation performance. The number of pixels with elevation ≤4000 m =28 whereas the number of pixels with elevation >4000 m = 27.

B. SWE evaluation

Fig. 7 displays the OL and DA SWE estimates for two test sites. The test sites #1 and #2 included in Fig. 7 are the same as those presented in Fig. 6. The number of grid cells included for SWE analysis in this section equals 38 based on the 55 station locations where some grid cells contain more than one AKAH station.

1) Standard assimilation: The DA SWE estimates (Fig. 7a) resemble the snow depth timeseries (Fig. 6) with better DA and SNOWPACK-derived SWE conformity (relative to the open loop) during the snow accumulation period, especially for test site #1. For test site #2, the complete melt-off date occurs earlier than the OL snowpack melt-off and is closer to the SNOWPACK-derived SWE estimates. For test site #1, the large DA SWE updates during Apr and May appear erroneous as the SWE values change ∼50 cm overnight. These large updates are a consequence of the ill-posed nature of PMW remote sensing of snow when the snowpack is wet coupled with controllability issues in the SVM-based observation operators in the presence of snow liquid water content [14].

2) Data thinning-based assimilation: The saw-tooth pattern and the incorporation of high frequency noise (information not related to snow mass) motivates the exploration of data thinning. Fig. 7b shows the effect of data thinning on DA SWE estimates. Test sites #1 and #2 exhibit a similar temporal pattern in terms of better consistency with SNOWPACK SWE during the snow accumulation period and large SWE updates during the ablation period. The corresponding ensemble uncertainty is also visibly increased during Apr and May. Similar to the snow depth timeseries (Fig. 6b), the data thinning experiment yields SWE estimates that exhibit a temporal pattern of smooth snowpack development for short intervals with visible SWE updates via assimilation on incremental days.

3) Seasonal assimilation: Test site #1 in Fig. 7c shows improved results for the seasonal assimilation run as the snowpack melt-off date more closely agrees with SNOWPACK

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3102965, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

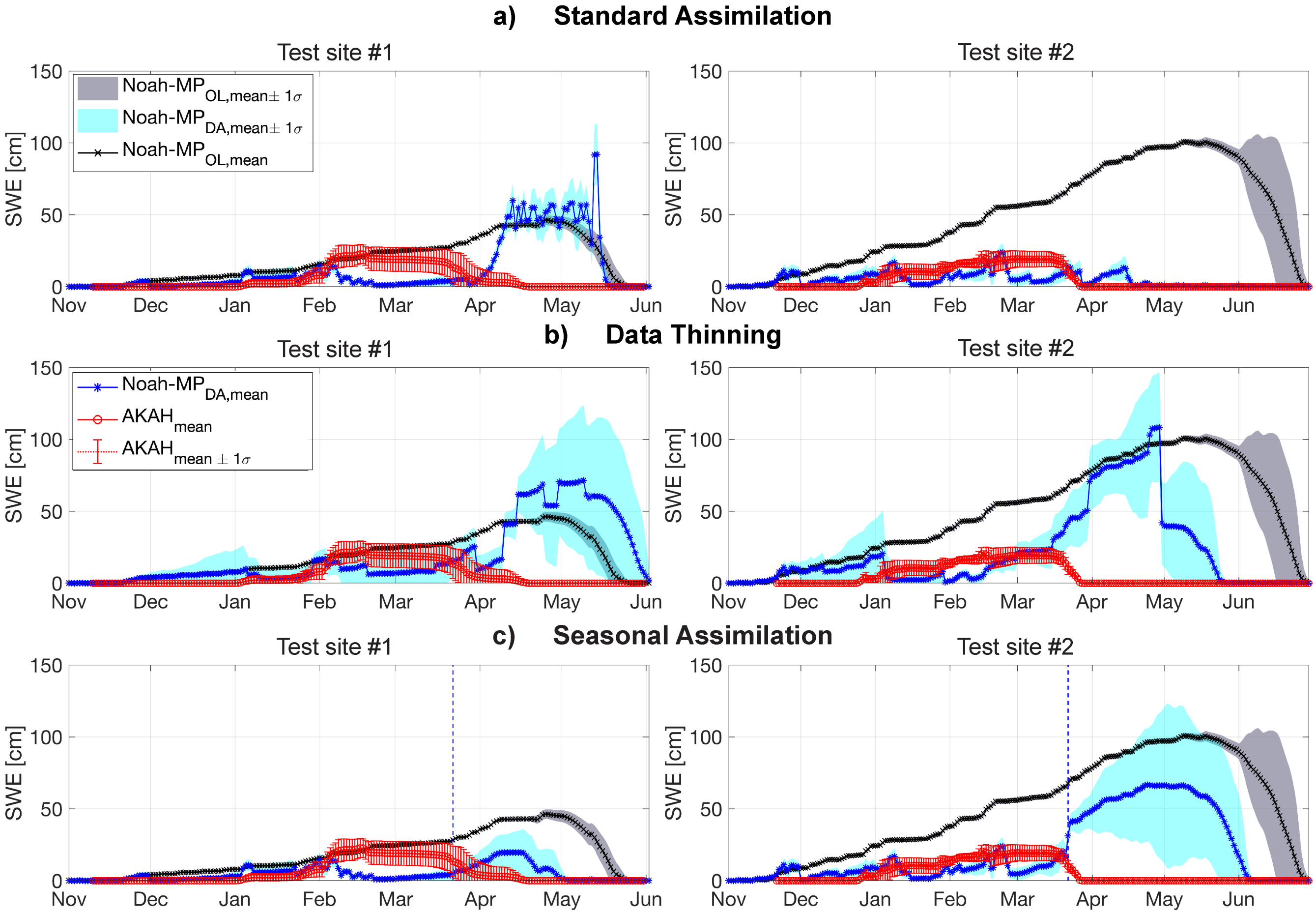

Fig. 7: Comparative timeseries of OL and DA SWE estimates vs. SNOWPACK SWE (modeled using AKAH snow depth measurements) for the 2016-2017 snow season. The dashed lines in part c) locate the date (22 March) when assimilation is disabled for seasonal assimilation. The red solid line represents the mean while the red whiskers represent the standard deviation calculated from SNOWPACK SWE estimates for stations that lie within the Noah-MP grid cell.

SWE. Test site #2 shows improved estimates during the accumulation period, but a large SWE update immediately before the assimilation is disabled (22 March 2017) considerably increases the ensemble uncertainty and the ensemble mean magnitude. For test site #2, the estimates are similar to the OL here onward as the ensemble is propagated forward in time without further assimilation.

4) Statistical results: Table VI contains the domain-averaged statistics for the three different assimilation runs. The domain-averaged bias is improved for all of the assimilation runs as compared to the OL with the highest decrease in domain-averaged bias obtained via standard assimilation. Data thinning-based estimates exhibit relatively higher domain-averaged bias and RMSE as compared to the other assimilation runs (Table VI). The domain-averaged bias for seasonal assimilation is comparatively on the lower end. The lowest domain-averaged and domain-median RMSE are exhibited by the seasonal assimilation approach. It must be noted that these are domain-averaged values that can be significantly influenced by large positive or negative bias values at a few locations.

Analyzing the computed statistics with respect to land cover type and elevation, it was noted that, similar to snow depth,

Statistic [cm]	Open Loop	Standard Assimilation	Data Thinning	Seasonal Assimilation
Mean Bias	24.2	14.2	20.4	14.3
Median Bias	19.6	8.7	13.7	9.4
Mean RMSE	34.1	33.6	37.0	29.2
Median RMSE	27.2	24.2	29.6	20.4
Mean Unbiased RMSE	22.8	27.6	28.5	23.0
Median Unbiased RMSE	19.8	22.8	24.4	17.5

TABLE VI: Domain-averaged and domain-median bias and RMSE of OL and DA SWE estimates as compared to the SNOWPACK SWE values for the 2016-17 snow season. All values are in units of centimeter.

Run Type	SWE			
	Elevation of Noah-MP pixels			
	Mean Bias [cm]		Mean RMSE [cm]	
	≤4000 m	>4000 m	≤4000 m	>4000 m
Open Loop	30.5	14.5	40.3	24.6
Standard Assimilation	15.5	12.3	35.4	30.8
Data Thinning	24.0	15.0	40.7	31.3
Seasonal Assimilation	15.9	11.7	30.6	27.1

TABLE VII: Influence of elevation on ΔT_b assimilation performance. The number of pixels with elevation ≤4000 m = 23 whereas the number of pixels with elevation >4000 m = 15.

the average bias in SWE estimates for pixels categorized as barren land cover was relatively lower for standard and seasonal assimilation runs (11.6 cm and 9.0 cm respectively) as compared to the OL and data thinning runs (25.0 cm and 23.6 cm respectively). Correlations calculated between pixel elevations and statistics (bias and RMSE) were; i) all negative (i.e., as elevation increases the SWE estimates improve), ii) $|R|$ <0.5 for all the DA runs, and iii) equal to -0.5 for the OL. Table VII presents the influence of assimilation on statistics calculated for pixels within different elevation ranges. For pixels at elevation ≤4000 m, standard and seasonal assimilation show relatively lower mean bias and RMSE as compared to the OL. For elevation >4000 m, where deep snow is expected, the OL exhibits a lower mean RMSE than the assimilation runs.

C. Snow accumulation period

Figs. 6 and 7 highlighted that the DA estimates performed much better during the snow accumulation period as compared to the ablation period. The improved performance is due to the relatively higher presence of snow mass-relevant information and the reduction of background noise (i.e., information related to features other than snow mass) in the AMSR2 ΔT_b signal during the snow accumulation period. Figs. 8 and 9 show the maps of snow depth and SWE statistics, respectively, computed for the snow accumulation months only (Nov-Feb). The standard assimilation run is used for both the snow depth and the SWE comparison.

The domain-averaged bias and RMSE for both snow depth and SWE for the OL and the DA are visibly reduced as compared to the complete snow season statistics, presented in Tables IV and VI. This is due to the temporal lag in the snowpack development exhibited by the OL and the DA runs as compared to the AKAH measurements, i.e., the OL and DA simulation peak value occurs after February thus the snow depth and SWE magnitudes of all the estimates are much lower resulting in a smaller bias and RMSE magnitude. The DA snow depth estimates (Fig. 8) exhibit reduced domain-averaged bias and RMSE values as compared to the OL. The Noah-MP$_{OL}$ domain-averaged bias is reduced by 63% through

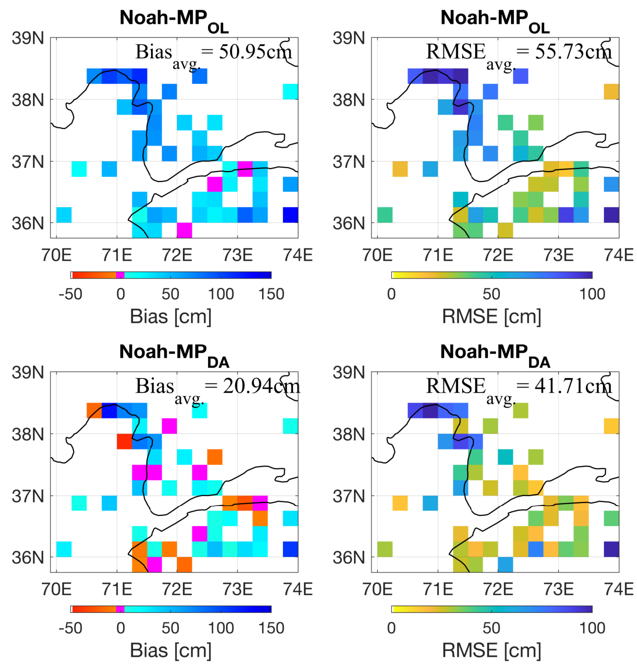

Fig. 8: OL and DA (standard assimilation) maps of bias and RMSE for snow depth estimates during the snow accumulation months (Nov-Feb), 2016-17.

ΔT_b assimilation. The snow depth absolute bias and RMSE were reduced at 47 and 42 (out of 55 total) comparison cells, respectively. Fig. 9 shows a similar reduction in DA SWE bias and RMSE. The OL and DA RMSE maps show similar domain-averaged RMSE magnitudes. The absolute bias and RMSE for the SWE estimates were reduced in magnitude at 29 and 22 (out of 38 total) comparison cells, respectively, as a result of ΔT_b assimilation when considering only the snow accumulation period.

VI. Discussion

According to Tables IV and VI, daily assimilation performed better than incremental-day (data thinning) assimilation as both standard and seasonal assimilation have lower domain-averaged mean and domain-averaged median bias and RMSE as compared to the data thinning assimilation run. Data thinning was performed in an effort to extract snow mass-related information from the ΔT_b signal while reducing the import of high frequency noise and snow mass-irrelevant information into the updated timeseries. However, from Figs. 6b and 7b it is apparent that data thinning resulted in increased uncertainty within the ensemble and, in general, did not serve the intended goal. For seasonal assimilation, we used a first-order approach to identify wet snowpack conditions. Instead of relying on rainfall data, utilization of a more sophisticated approach may yield different results. Meanwhile, it is observed that the standard daily ΔT_b assimilation did reduce the bias and RMSE (as compared to the OL), especially during the snow accumulation months (Nov-Feb). The relatively better results observed during the (dry) snow accumulation season are expected. Several studies have analyzed the differences in

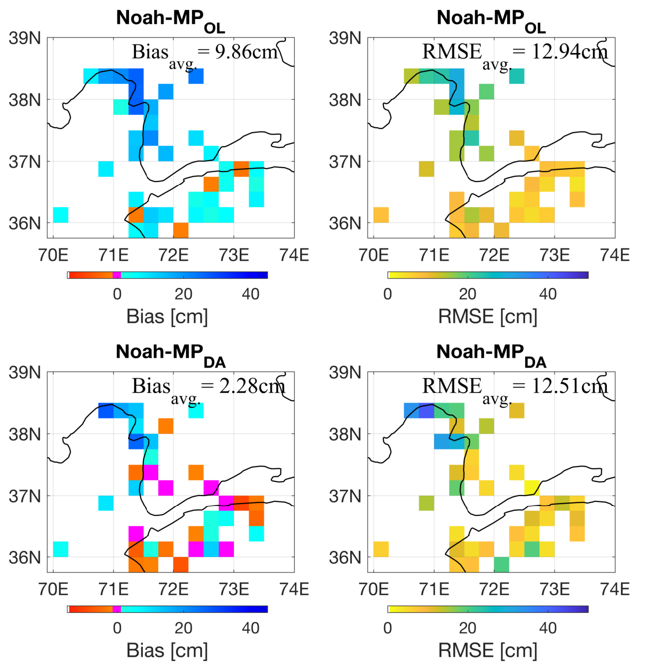

Fig. 9: OL and DA (standard assimilation) maps of bias and RMSE for SWE estimates during the snow accumulation months (Nov-Feb), 2016-17.

the relationship between SWE/SD and T_b for dry versus wet snow conditions [5], [50], [54]. Based on the conclusions of these studies, it is expected that the effectiveness of extracting SWE-relevant information from ΔT_b observations during the snow ablation period would be limited.

The results detailed in Section-V indicate the presence of a general positive bias for all SWE estimates. The positive bias highlights over-prediction of the OL and DA snow depth and SWE estimates in the evaluated sub-domain (AKAH domain). One reason behind the errors in SD and SWE estimates is the bias in the boundary conditions (i.e., MERRA2). The Noah-MP land surface model was used to estimate SD and SWE using MERRA2 boundary conditions (including precipitation). Through previous analysis, it was observed that in the complex, high-elevation terrain of HMA, MERRA2 exhibits certain biases [55]. High precipitation magnitudes are observed during the winter season and a lag is present in the onset of melt. As a result, the model predicts higher SD and SWE magnitudes for the snowpack, which completely melts at a later date as compared to the in situ data. A future trajectory for this study could be the replication of the described experiments using alternative model boundary conditions. Another possible cause of error is the Noah-MP land surface model's formulation of snow albedo– snow albedo influences the development and melting of the snowpack. Kumar et al. [56] studied the spectral components of albedo estimated by Noah-MP and concluded that errors in the visible albedo components affected the reflective radiative fluxes, and hence, the subsequent development and melting of the snowpack.

The general, positive bias could also be due to the elevation difference between the AKAH stations and the Noah-MP grid cells. The Noah-MP grid cells are generally at a higher elevation as compared to the corresponding AKAH stations, and therefore, the snowpack peak SD/SWE and total meltout in Noah-MP occur at a later date. It should also be stated that the comparisons detailed above are an evaluation, not a validation, of the OL and the DA results as: i) differences in the spatial scales of the in situ measurements versus the Noah-MP output can give rise to representativeness error, and ii) the evaluation SWE data is also modeled, although it is predicated on the in situ AKAH snow depth measurements.

The results included in Section-V are computed by comparing the OL and DA estimates with the AKAH measurements. These measurements were taken at stations located in the northwestern part of HMA. Previous studies [4], [21] have established the presence of different hydrologic regimes operating within different parts of HMA. Therefore, the results presented here and the conclusions drawn from them are applicable primarily to the northwestern part of HMA.

A limiting feature in the assimilation framework presented here is the SVM-based observation operator. It limits the applicability of the assimilation framework at certain instances in time, especially during the ablation season, when the sensitivity of trained SVMs to input SWE is reduced [42]. In addition, information within the AMSR2 ΔT_b observations related to physical features not related to snow mass (and which subsequently acts as noise in this framework) is also transferred to the trained SVMs and can result in erroneous ΔT_b training and prediction. The ill-posed nature of ΔT_b translation to SWE in the presence of wet snowpacks is further heightened when attempting this conversion in complex terrain.

VII. Conclusion

In this study, a PMW ΔT_b assimilation framework was utilized to improve snow estimates over high mountain Asia. Well-trained SVMs served as the observation operators. Snow depth and SWE modeled via SNOWPACK based on boundary conditions specified by in situ snow depth measurements were used during evaluation in order to asses the accuracy of the Noah-MP model only (OL) snow estimates versus the ΔT_b assimilated (DA) estimates. The computed statistics highlight the improvements in domain-averaged performance of the DA estimates. Absolute bias in SWE estimates was improved at 74% while RMSE was reduced at 68% of the total comparative pixels through standard assimilation. Comparing the bias and RMSE results, it can be stated that assimilation proved more effective in reducing bias than RMSE.

SVM training is an important component of the outlined framework and it is thus necessary that the SVM training (input and target) data be analyzed prior to the training process. The ill-posed nature of the ΔT_b to SWE conversion problem is partially responsible for the degradation of the estimation accuracy at particular instances in space and time, especially during the ablation period when the AMSR2 ΔT_b observations contain considerable amount of information not relevant to snow mass thus reducing the relevant information content

in the PMW signal. The relatively improved performance of the described framework during the (dry) snow accumulation period and weak performance during the (wet) snow ablation period is expected considering the results of relevant previous studies [5], [50].

The experimental results detailed above show that the AMSR2 passive microwave brightness temperatures do add utility to modeled SWE estimates in complex terrain, specifically during the dry snow accumulation months. The coarse PMW ΔT_b observations can be used in areas where finer resolution data does not exist in order to achieve improvements in snow depth and SWE estimation. This study attempts to fill the gap in our knowledge regarding consistent snow mass estimation in HMA, particularly in the northwestern sub-region, through the use of PMW ΔT_b. Although the utilization of PMW for accurate snow estimation in complex terrain is dependent on a multitude of factors, it does exhibit certain utility for improving snow estimation in parts of HMA. One potential path forward is the development and assimilation of physically-enhanced brightness temperature combinations as compared to the simple ΔT_b combinations assimilated in this study. An example would be the utilization of an approach similar to that described by Takala et al. [57] who used a forest factor to scale the ΔT_b magnitude. Another potential strategy could be based on using a cost-function to control the magnitude of SWE update during the wet ablation season using some form of a variational assimilation approach.

Appendix A
Support Vector Machine Regression Theory

The Vapnik-Chervonenkis theory [39] forms the basis of SVM regression. In ε - SV regression [58], the objective is to find a function *f(x)* that has, at most, ε deviations from the training targets, y_i. That is, ε defines the value of the permissible error. A linear form of such an objective function is:

$$f(x) = \langle w, x \rangle + b \quad \text{such that } b \in \mathbb{R} \tag{6}$$

where $\langle \, . \, , \, . \, \rangle$ is the dot product in X (X is the input pattern space and contains the vector x), w is a vector of weights, and $\mathbb{R}$ represents the real number space. This problem can be formulated as an optimization problem [40] as:

$$\begin{aligned} &\text{minimize} && \frac{1}{2}\|w\|^2 \\ &\text{subject to} && \begin{cases} y_i - \langle w, x \rangle - b \leq \varepsilon \\ \langle w, x \rangle + b - y_i \leq \varepsilon \end{cases} \end{aligned} \tag{7}$$

where $\|.\|$ represents the Euclidean norm.

A non-linear version of this objective function uses a kernel function to map the input data into a higher dimensional feature space where the data displays a linear relationship. The radial basis function (RBF) kernel was used in this study. The RBF kernel has one parameter γ which describes its bandwidth.

The SVM algorithm was implemented using the LIBSVM library provided by National Taiwan University [59]. The ε - SV regression requires user-provided values for three parameters: $i)$ C, $ii)$ ε, and $iii)$ γ. C is a penalty parameter and is set as the range of the training targets (y_i) in this study. ε and γ were selected using a two-phase cross-validation method [41], [42]. A separate SVM was trained for each grid cell for each fortnight. Each fortnightly SVM was trained using data from the relevant fortnight as well as two weeks prior to and two weeks after the fortnight under consideration in order to improve temporal continuity from one fortnight to the next. Thus, each SVM was trained using a six-week input dataset and a total of 26 SVMs were produced for each year at each location in space.

Appendix B
Statistical Formula

The following formulas were used to calculate the bias and RMSE, respectively:

$$Bias = \sum_{t=1}^{T} (y_{\text{simulated}} - y_{\text{observed}}) \tag{8}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T} (y_{\text{simulated}} - y_{\text{observed}})^2}{T}} \tag{9}$$

where $y_{\text{simulated}}$ equals the OL/DA snow depth or SWE estimate (ensemble mean), y_{observed} is the AKAH snow depth or SNOWPACK SWE, and T is the total number of data instances in time at a given location in space.

Appendix C
SNOWPACK

For a more detailed description of the SNOWPACK model runs please see Bair et al. [60] such that only a summary is provided here. SNOWPACK v3.5 was run in research mode at a 15 minute time step with hourly outputs for each of the AKAH stations. Hourly forcings were computed by combining temporally interpolated snow depth from the AKAH in situ measurements with air temperature, incoming shortwave, reflected shortwave, incoming longwave, wind speed, and relative humidity. Radiative forcings were obtained from CERES Ed 4a [49] whereas all the other forcings were collected from GLDAS-2 [48]. These inputs were then downscaled spatially using ParBal [61]. SNOWPACK was only run for periods when measurements from the AKAH stations were available, i.e., Nov/Dec to Apr/May, depending on the year. Grid cells were assumed to be level, hence, forcings without terrain correction were applied except for shading effects when the Sun was below the local horizon [62]. The wind direction, which is not available in GLDAS-2, was fixed as the mean value from the daily AKAH instantaneous values. The ground temperature was set as the daily minimum of the air temperature or -1.5°C when snow cover was present. Changes to default parameters that affected model output are provided in Table A2 of Bair et al. [60].

Acknowledgment

Funding for J.A., B.F., and S.K. was provided by the NASA High Mountain Asia Science Team (NNX17AC15G), while E.B. was funded through NASA contracts 80NSSC18K0427 and 80NSSC18K1489. We thank the Aga Khan Agency for Habitat and Doug Chabot for sharing snow measurements from the Pakistan-Afghanistan-Tajikistan region.

Data Statement

The source code for NASA Land Information System is available on GitHub (*https://github.com/NASA-LIS/LISF.git*). The code for SNOWPACK is accessible at *https://models.slf.ch/p/snowpack/*. The GLDAS-2 (*https://doi.org/10.5067/L0JGCNVBNRAX*, [48]) and MERRA-2 (*https://doi.org/10.5067/VJAFPLI1CSIV*, [27]) forcings are accessible at *https://disc.gsfc.nasa.gov/*. Due to the sensitive geographical location of the study domain, the snow depth measurements used in this study are not publicly available. Requests for the dataset should be made through the Aga Khan Agency for Habitat (*https://www.akdn.org*).

References

[1] W. W. Immerzeel, L. P. H. van Beek, and M. F. P. Bierkens, "Climate change will affect the Asian water towers," *Science*, vol. 328, no. 5984, pp. 1382–1385, 2010.

[2] A. Lutz, W. Immerzeel, A. Shrestha, and M. Bierkens, "Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation," *Nature Climate Change*, vol. 4, no. 7, p. 587, 2014.

[3] J. Xu, R. E. Grumbine, A. Shrestha, M. Eriksson, X. Yang, Y. Wang, and A. Wilkes, "The melting Himalayas: cascading effects of climate change on water, biodiversity, and livelihoods," *Conservation Biology*, vol. 23, no. 3, pp. 520–530, 2009.

[4] P. Wester, A. Mishra, A. Mukherji, and A. B. Shrestha, *The Hindu Kush Himalaya Assessment- Mountains, Climate Change, Sustainability and People*. Springer Nature Switzerland, 2018.

[5] A. Chang, J. Foster, D. Hall, A. Rango, and B. Hartline, "Snow water equivalent estimation by microwave radiometry," *Cold Regions Science and Technology*, vol. 5, no. 3, pp. 259–267, 1982.

[6] P. R. Singh and T. Y. Gan, "Retrieval of snow water equivalent using passive microwave brightness temperature data," *Remote Sensing of Environment*, vol. 74, no. 2, pp. 275–286, 2000.

[7] H. J. Zwally and P. Gloersen, "Passive microwave images of the polar regions and research applications," *Polar Record*, vol. 18, no. 116, pp. 431–450, 1977.

[8] R. Kelly, "The amsr-e snow depth algorithm: Description and initial results," *Journal of the Remote Sensing Society of Japan*, vol. 29, no. 1, pp. 307–317, 2009.

[9] K. Rittger, E. H. Bair, A. Kahl, and J. Dozier, "Spatial estimates of snow water equivalent from reconstruction," *Advances in Water Resources*, vol. 94, pp. 345–363, 2016.

[10] N. P. Molotch and S. A. Margulis, "Estimating the distribution of snow water equivalent using remotely sensed snow cover data and a spatially distributed snowmelt model: A multi-resolution, multi-sensor comparison," *Advances in Water Resources*, vol. 31, no. 11, pp. 1503 – 1514, 2008, hydrologic Remote Sensing.

[11] J. D. Kirkham, I. Koch, T. M. Saloranta, M. Litt, E. E. Stigter, K. MÃ¸en, A. Thapa, K. Melvold, and W. W. Immerzeel, "Near real-time measurement of snow water equivalent in the Nepal Himalayas," *Frontiers in Earth Science*, vol. 7, p. 177, 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/feart.2019.00177

[12] E. H. Bair, A. Abreu Calfa, K. Rittger, and J. Dozier, "Using machine learning for real-time estimates of snow water equivalent in the watersheds of Afghanistan." *Cryosphere*, vol. 12, no. 5, 2018.

[13] E. E. Stigter, N. Wanders, T. M. Saloranta, J. M. Shea, M. F. Bierkens, and W. W. Immerzeel, "Assimilation of snow cover and snow depth into a snow model to estimate snow water equivalent and snowmelt runoff in a Himalayan catchment," *The Cryosphere*, vol. 11, no. 4, pp. 1647–1664, 2017.

[14] Y. Kwon, B. A. Forman, J. A. Ahmad, S. V. Kumar, and Y. Yoon, "Exploring the utility of machine learning-based passive microwave brightness temperature data assimilation over terrestrial snow in high mountain asia," *Remote Sensing*, vol. 11, no. 19, p. 2265, 2019.

[15] M. Sturm, J. Holmgren, and G. E. Liston, "A seasonal snow cover classification system for local to global applications," *Journal of Climate*, vol. 8, no. 5, pp. 1261–1283, 1995.

[16] J. C. Hammond, F. A. Saavedra, and S. K. Kampf, "Global snow zone maps and trends in snow persistence 2001–2016," *International Journal of Climatology*, vol. 38, no. 12, pp. 4369–4383, 2018.

[17] P. Singh and N. Kumar, "Impact assessment of climate change on the hydrological response of a snow and glacier melt runoff dominated Himalayan river," *Journal of Hydrology*, vol. 193, no. 1, pp. 316 – 350, 1997.

[18] K. Rühland, N. R. Phadtare, R. K. Pant, and S. J. Sangode, "Accelerated melting of Himalayan snow and ice triggers pronounced changes in a valley peatland from northern India," *Geophysical Research Letters*, vol. 33, no. 15, 2006.

[19] M. Adnan, G. Nabi, M. S. Poomee, and A. Ashraf, "Snowmelt runoff prediction under changing climate in the Himalayan cryosphere: A case of Gilgit river basin," *Geoscience Frontiers*, vol. 8, no. 5, pp. 941 – 949, 2017.

[20] T. Smith and B. Bookhagen, "Changes in seasonal snow water equivalent distribution in high mountain Asia (1987 to 2009)," *Science advances*, vol. 4, no. 1, p. e1701550, 2018.

[21] R. L. Armstrong, K. Rittger, M. J. Brodzik, A. Racoviteanu, A. P. Barrett, S.-J. S. Khalsa, B. Raup, A. F. Hill, A. L. Khan, A. M. Wilson *et al.*, "Runoff from glacier ice and seasonal snow in high asia: separating melt water sources in river flow," *Regional Environmental Change*, pp. 1–13, 2018.

[22] J. S. Kargel, G. J. Leonard, M. P. Bishop, A. Kääb, and B. H. Raup, *Global land ice measurements from space*. Springer, 2014.

[23] S. V. Kumar, C. D. Peters-Lidard, Y. Tian, P. R. Houser, J. Geiger, S. Olden, L. Lighty, J. L. Eastman, B. Doty, P. Dirmeyer *et al.*, "Land information system: An interoperable framework for high resolution land surface modeling," *Environmental modelling & software*, vol. 21, no. 10, pp. 1402–1415, 2006.

[24] M. Ek, K. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. Tarpley, "Implementation of noah land surface model advances in the national centers for environmental prediction operational mesoscale eta model," *Journal of Geophysical Research: Atmospheres*, vol. 108, no. D22, 2003.

[25] G.-Y. Niu, Z.-L. Yang, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, A. Kumar, K. Manning, D. Niyogi, E. Rosero *et al.*, "The community noah land surface model with multiparameterization options (noah-mp): 1. model description and evaluation with local-scale measurements," *Journal of Geophysical Research: Atmospheres*, vol. 116, no. D12, 2011.

[26] Z.-L. Yang, G.-Y. Niu, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, L. Longuevergne, K. Manning, D. Niyogi, M. Tewari *et al.*, "The community noah land surface model with multiparameterization options (noah-mp): 2. evaluation over global river basins," *Journal of Geophysical Research: Atmospheres*, vol. 116, no. D12, 2011.

[27] R. Gelaro, W. McCarty, M. J. Suárez, R. Todling, A. Molod, L. Takacs, C. A. Randles, A. Darmenov, M. G. Bosilovich, R. Reichle *et al.*, "The modern-era retrospective analysis for research and applications, version 2 (merra-2)," *Journal of Climate*, vol. 30, no. 14, pp. 5419–5454, 2017.

[28] K. R. Arsenault, S. V. Kumar, J. V. Geiger, S. Wang, E. Kemp, D. M. Mocko, H. K. Beaudoing, A. Getirana, M. Navari, B. Li *et al.*, "The Land surface Data Toolkit (LDT v7.2)–a data fusion environment for land data assimilation systems," *Geoscientific Model Development*, vol. 11, no. 9, pp. 3605–3621, 2018.

[29] T. G. Farr, P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth *et al.*, "The shuttle radar topography mission," *Reviews of geophysics*, vol. 45, no. 2, 2007.

[30] M. A. Friedl, D. K. McIver, J. C. Hodges, X. Y. Zhang, D. Muchoney, A. H. Strahler, C. E. Woodcock, S. Gopal, A. Schneider, A. Cooper *et al.*, "Global land cover mapping from modis: algorithms and early results," *Remote sensing of Environment*, vol. 83, no. 1-2, pp. 287–302, 2002.

[31] D. A. Robinson and G. Kukla, "Maximum surface albedo of seasonally snow-covered lands in the Northern Hemisphere," *Journal of Climate and Applied Meteorology*, vol. 24, no. 5, pp. 402–411, 1985.

[32] G. Gutman and A. Ignatov, "The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models," *International Journal of remote sensing*, vol. 19, no. 8, pp. 1533–1543, 1998.

[33] J. T. Ball, I. E. Woodrow, and J. A. Berry, "A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions," in *Progress in photosynthesis research*. Springer, 1987, pp. 221–224.

[34] G.-Y. Niu, Z.-L. Yang, R. E. Dickinson, L. E. Gulden, and H. Su, "Development of a simple groundwater model for use in climate models and evaluation with gravity recovery and climate experiment data," *Journal of Geophysical Research: Atmospheres*, vol. 112, no. D7, 2007.

[35] G.-Y. Niu and Z.-L. Yang, "Effects of frozen soil on snowmelt runoff and soil water storage at a continental scale," *Journal of Hydrometeorology*, vol. 7, no. 5, pp. 937–952, 2006.

[36] W. Brutsaert, *Evaporation into the atmosphere: theory, history and applications*. Springer Science & Business Media, 2013, vol. 1.

[37] Z.-L. Yang and R. E. Dickinson, "Description of the Biosphere-Atmosphere Transfer Scheme (BATS) for the soil moisture workshop and evaluation of its performance," *Global and Planetary Change*, vol. 13, no. 1-4, pp. 117–134, 1996.

[38] R. Jordan, "A one-dimensional temperature model for a snow cover: Technical documentation for sntherm. 89." Cold Regions Research and Engineering Lab Hanover, NH, Tech. Rep., 1991.

[39] V. N. Vapnik, A. Y. Chervonenkis, and N. Moskva, "Pattern recognition theory," *Statistical Learning Problems*, 1974.

[40] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.

[41] B. A. Forman and R. H. Reichle, "Using a support vector machine and a land surface model to estimate large-scale passive microwave brightness temperatures over snow-covered land in North America," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 9, pp. 4431–4441, 2014.

[42] J. A. Ahmad, B. A. Forman, and Y. Kwon, "Analyzing machine learning predictions of passive microwave brightness temperature spectral difference over snow-covered terrain in high mountain asia," *Frontiers in Earth Science*, vol. 7, p. 212, 2019.

[43] Y. Xue, B. A. Forman, and R. H. Reichle, "Estimating snow mass in North America through assimilation of AMSR-E brightness temperature observations using the catchment land surface model and support vector machines," *Water resources research*, vol. 54, no. 9, p. 6488, 2018.

[44] C. Mätzler, "Applications of the interaction of microwaves with the natural snow cover," *Remote sensing reviews*, vol. 2, no. 2, pp. 259–387, 1987.

[45] M. Lehning, P. Bartelt, B. Brown, C. Fierz, and P. Satyawali, "A physical snowpack model for the swiss avalanche warning: Part ii. snow microstructure," *Cold regions science and technology*, vol. 35, no. 3, pp. 147–167, 2002a.

[46] M. Lehning, P. Bartelt, B. Brown, and C. Fierz, "A physical snowpack model for the swiss avalanche warning: Part iii: Meteorological forcing, thin layer formation and evaluation," *Cold Regions Science and Technology*, vol. 35, no. 3, pp. 169–184, 2002b.

[47] E. H. Bair, R. E. Davis, and J. Dozier, "Hourly mass and snow energy balance measurements from mammoth mountain, ca usa, 2011–2017," *Earth System Science Data*, vol. 10, no. 1, pp. 549–563, 2018.

[48] M. Rodell, P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C. J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J. K. Entin, J. P. Walker, D. Lohmann, and D. Toll, "The global land data assimilation system," *Bulletin of the American Meteorological Society*, vol. 85, no. 3, pp. 381–394, 2004.

[49] D. A. Rutan, S. Kato, D. R. Doelling, F. G. Rose, L. T. Nguyen, T. E. Caldwell, and N. G. Loeb, "Ceres synoptic product: Methodology and validation of surface radiant flux," *Journal of Atmospheric and Oceanic Technology*, vol. 32, no. 6, pp. 1121–1143, 2015.

[50] A. D. Frolov and Y. Ya. Macheret, "On dielectric properties of dry and wet snow," *Hydrological processes*, vol. 13, no. 12-13, pp. 1755–1760, 1999.

[51] J. Wang, B. A. Forman, and Y. Xue, "Exploration of synthetic terrestrial snow mass estimation via assimilation of amsr-e brightness temperature spectral differences using the catchment land surface model and support vector machine regression," *Water Resources Research*, p. e2020WR027490, 2020.

[52] M. Durand and S. A. Margulis, "Feasibility test of multifrequency radiometric data assimilation to estimate snow water equivalent," *Journal of Hydrometeorology*, vol. 7, no. 3, pp. 443–457, 2006.

[53] NIC, "IMS daily northern hemisphere snow and ice analysis at 4 km and 24 km resolution, version 1," 2008. [Online]. Available: https://nsidc.org/data/

[54] A. Rango, "Operational applications of satellite snow cover observations," *JAWRA Journal of the American Water Resources Association*, vol. 16, no. 6, pp. 1066–1073, 1980.

[55] Y. Yoon, S. V. Kumar, B. A. Forman, B. F. Zaitchik, Y. Kwon, Y. Qian, S. Rupper, V. Maggioni, P. Houser, D. Kirschbaum *et al.*, "Evaluating the uncertainty of terrestrial water budget components over high mountain asia," *Frontiers in Earth Science*, vol. 7, p. 120, 2019.

[56] S. Kumar, D. Mocko, C. Vuyovich, and C. Peters-Lidard, "Impact of surface albedo assimilation on snow estimation," *Remote Sensing*, vol. 12, no. 4, p. 645, 2020.

[57] M. Takala, K. Luojus, J. Pulliainen, C. Derksen, J. Lemmetyinen, J.-P. Kärnä, J. Koskinen, and B. Bojkov, "Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements," *Remote Sensing of Environment*, vol. 115, no. 12, pp. 3517–3529, 2011.

[58] V. N. Vapnik, *The nature of statistical learning theory*. Springer, 1995.

[59] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[60] E. H. Bair, K. Rittger, J. A. Ahmad, and D. Chabot, "Comparison of modeled snow properties in afghanistan, pakistan, and tajikistan." *Cryosphere*, vol. 14, no. 1, 2020.

[61] E. H. Bair, K. Rittger, R. E. Davis, T. H. Painter, and J. Dozier, "Validating reconstruction of snow water equivalent in California's Sierra Nevada using measurements from the NASA Airborne Snow Observatory," *Water Resources Research*, vol. 52, no. 11, pp. 8437–8460, 2016.

[62] J. Dozier and J. Frew, "Rapid calculation of terrain parameters for radiation modeling from digital elevation data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 5, pp. 963–969, 1990.

Jawairia A. Ahmad received her M.S. degree in Civil Engineering from the Department of Civil and Environmental Engineering, University of Maryland, College Park, USA, where she is currently pursuing a Ph.D. Her doctoral dissertation is focused on improving terrestrial water budget estimation across high mountain Asia and the adjoining region. Her research interests include the utilization of machine learning and data assimilation in Earth science.

Barton A. Forman received a B.S. (*cum laude*) degree in civil engineering from the University of Virginia, Charlottesville, VA, a M.S. degree in civil and environmental engineering from the University of California at Berkeley, Berkeley, CA, and a Ph.D. degree in civil engineering from the University of California at Los Angeles, Los Angeles, CA. Currently, he is an Associate Professor with the Department of Civil and Environmental Engineering, University of Maryland, College Park, MD. His research interests include estimation of snowpack using space-based measurements of the Earth's gravitational field as well as high-performance computing applications related to terrestrial hydrology.

Edward (Ned) H. Bair received his B.A. in economics (magna cum laude) from Bowdoin College in 2003 and his Ph.D. in Environmental Science and Management in 2011 from the University of California, Santa Barbara. He has been a member of the research faculty at UC Santa Barbara since 2011. He uses remote sensing and field techniques to study the snowpack. He specializes in snow mapping, energy balance modeling, and avalanche formation in montane regions. His work spans a range of scales, from a few meters at the plot scale to thousands of square kilometers at the basin and mountain range scale. He has over 50 publications on snow hydrology and snow avalanches. He has worked for the US Army Corps of Engineers as a consultant. Dr. Bair is a member of the American Geophysical Union and a professional member of the American Avalanche Association and its Research Chair. In 2015, he received the Best Oral Paper Award from the Western Snow Conference.

Sujay V. Kumar received Bachelor of Technology from the Indian Institute of Technology, Bombay in Civil Engineering in 1996, and Masters (1998) and a Ph.D. degree (2002) in Computer Aided Engineering from North Carolina State University, Raleigh, NC. He is currently a research physical scientist at the Hydrological Sciences Lab at the NASA Goddard Space Flight Center. He is the lead of the NASA Land Information System (LIS) team, a widely-used high performance platform for land surface modeling and data assimilation. His research interests include land surface modeling, data assimilation, remote sensing, high-performance computing, and machine learning.